WIKTOR PACIOREK, JOANNA RĄCZASZEK-LEONARDI
University of Warsaw

THE INFLUENCE OF SENTENTIAL CONTEXT AND FREQUENCY OF OCCURRENCE ON THE RECOGNITION OF WORDS WITH SCRAMBLED LETTERS

In this paper we examine the "jumbled words" effect which denotes human ability to easily read words whose internal letters have been re-arranged as long as external letters remain in their positions. Hitherto, many explanations for this effect have focussed on the processes that operate "bottom-up". Here we suggest that "top-down" processes also play an important role and demonstrate this experimentally. First, we briefly describe the main types of word-recognition models and consider which model best explains the effect. Then, we present an experiment in which jumbled words of different frequency of occurrence were immersed in various types of contexts. Results indicate that both the frequency and semantic sentential context are involved in jumbled word recognition. The implications of these findings for word recognition models are discussed.

Key words: word recognition, "jumbled words", "top-down" processing

Introduction

Models of printed words' recognition may be divided into models stressing the "bottom-up" processing and models which take into account both the "bottom-up" and "top-down" direction of processing. The assumption for the "bottom-up" models is that the whole information necessary to recognize a word is contained in the stimulus. The cognitive system's "task" is to transform this information and to match it to a word in the mental dictionary. Any contextual information, if used at all, is utilized after the initial, automatic and context-free recognition. One of the arguments cited to support "bottom-up" models is that the process of

Address for correspondence: Wiktor Paciorek, University of Warsaw, Faculty of Psychology, Stawki 5/7, 00-183 Warsaw, Poland. E-mail: w.paciorek@gmail.com

recognizing words must be independent of "top-down" processing, because even people fluent in a certain language are able to predict only one in four words in a printed, logical and properly formulated text (Gough, Alford, Holley-Wilcox, 1981, after: Gleason, Ratner, 2005). However, it does not seem to constitute sufficient evidence, since the ability to guess one in four words in a text can be used as an argument for just the opposite statement: that the "top" information is utilized in the process of reading to a large extent. This is because the probability of word occurrence, without considering the semantic and/or syntactic context, is not higher than 1:100000.

The authors of models which acknowledge also the "top-down" factors suggest that, in recognizing words, apart from "bottom-up" processing, the previously stored information which makes the current context useful is used. The "top-down" models are sometimes described as context-driven word-recognition models. The word "context" usually refers to the semantic context, which is the information contained in the meaning of a sentence in which a word occurs. However, the semantic context is not the only indicator of "top-down" processing. The "top-down" information may also concern the word's frequency in language. The words which occur more frequently should be recognized more easily since their representations are also more easily activated (Whaley, 1978; Howes, Solomon, 1951, Grainger, 1990). The influence of the syntactic (Simpson, Peterson, Casteel, Burges, 1989), situational and social context may constitute other aspects of the "top-down" processing.

Models of word recognition and the "jumbled words" effect

The majority of word-recognition and fluent reading models deal with words belonging to a given language. However, it is difficult to find a model which would effectively explain the "jumbled words" effect, which refers to recognition of words with scrambled letters as actual ones. Most of the models proposed so far have been of the stimulus-driven character (Whitney, 2001; Coltheart et al., 2001). As stated earlier, such an approach seems sufficient for explaining the recognition of existing words. The models accounted for data both from experiments on reading by normal subjects, and from studies on changes in the process of reading in people with acquired post-traumatic dyslexia (Coltheart et al., 2001). However, the "jumbled words" effect is not easily accounted for in the "bottom-up" models. It has been claimed (Grainger and Whitney, 2004) that the difficulty might, for instance, stem from an incorrect mode of coding letters in words.

Many models hitherto proposed are based on the position-specific slot encoding. In this paradigm, each letter has many representations that take into account the position of a letter in a word. Thus there are many representations of a letter "A": as the first one in the word, as the second one, and so on. Nevertheless, the experiments of Humphreys (1990) and of Peressotti and Grainger (1990), which

concerned relative-position priming, have shown that the role of the exact order of letters is not as important as the letters' relative position. For example, the sequence of letters "tyki" may prime the word "kitty" because of the presence of letters "TY" and "KI" in each of them in the same order. This evidence provided the basis for Whitney's model of word recognition SERIOL, in which the letters are not encoded separately but in pairs, in the form of so-called open bigrams. Thus not only the exact order of letters in a word is taken into consideration, but rather their mutual relation. Grainger and Whitney (2004) have suggested that the "jumbled words" effect is best explained in SERIOL by open bigram encoding.

Another possible way of explaining the problem of the "jumbled words" effect is to focus on the "top-down" processing in word recognition (McClelland, Rumelhart, 1981). It is worth noticing that the issue of "top-down" processing has been repeatedly discussed; two general opinions emerge. The first, represented by a group of scientists called modularists, assumes that each word in a sentence constitutes a whole separated from other words and thus must be processed and recognized separately. Initially, the context is not taken into consideration (it may secondarily play an important role during attributing the meaning to words, for instance, in the case of ambiguous words). Such a stance is represented by Swinney (1979) and Fodor (1983). The second opinion is represented by interactionists (e.g. McClelland & Rumelhart, 1981; Tabossi, 1988). This view recognizes the importance of "top-down" processing, in particular the context, already at the moment of recognition of words in a sentence. McClelland and Rumelhart (1981) acknowledge various types of "top-down" factors and emphasize their interactions. For instance, if the effect of the sentential context is strong, the role of other factors (e.g. frequency of words in a language) might be smaller.

Some recent research (e.g. Velan & Frost, 2007) shows that the occurrence of the "jumbled words" effect might be language specific. In some languages, for example, Hebrew, a small change in letters (usually consonants), which constitute the meaningful core of the word, is likely to cause a large change in the meaning of the word. As a consequence, the "jumbled words" effect is much weaker in these languages, even if the context and frequency of the words used are similar to these in other languages. This possibly suggests that the "top-down" factors are influential, but only when the letter-derived processing remains undisturbed. Thus, to sufficiently explain the "jumbled words" effect, it might be crucial to find a word-recognition model which takes into account both "bottom-up" and "top-down" effects.

Research problem and hypothesis

The main goal of our experiments was to show that the recognition of "jumbled words" depends, aside from "bottom-up" stimulus properties, also on the "top-down" processing. "Jumbled words", created from words belonging to 2 frequency classes, were put into two kinds of sentence contexts. The dependent variable was the word-

reading time in a self-paced reading task. The hypothesis was that a semantically constraining context will shorten the time necessary for the recognition of a jumbled word, as compared to a semantically neutral context. What is more, it was expected that more frequent words would be recognized faster than infrequent words.

Material

Twenty nouns were selected for the experiment: 10 frequent and 10 infrequent ones. The frequency of words was determined on the basis of the PWN Polish Language Corpus (*Korpus języka polskiego PWN).* The infrequent words were those occurring 6-19 times in a million, the frequent ones occurred 84-173 times in a million. All words were either two or three syllables long.

In each word the internal letters were scrambled, while the first and last letters remained unchanged. It was also checked if the letters in both groups of words (infrequent and frequent) had been scrambled to a similar degree. This degree was determined by using the open-bigrams approach proposed by Whitney (2001), and described above. The quotient of shared bigrams of the word with scrambled letters and the appropriate word amounted to an average of 0.863 for infrequent words and 0.866 for frequent words.

For each word it was checked whether the sequence of letters created after scrambling does not constitute any other word existing in Polish, and that it is possible to "unscramble" the sequence in only one way. This operation was conducted in order to assure that jumbled-word recognition results only in the intended word.

Two context sentences were created for each word: a constraining one, in which the words presented before the stimulus word constrained semantically the possible words that could follow, and a neutral one, which did not provide such a constraint.

The extent to which a given sentence creates a constraining context was evaluated by three competent judges, fifth-year students of Polish philology. The judges received the beginnings of sentences to read. Their task was to suggest three words which, in their opinion, could occur after the introduction presented to them. If all three judges proposed as one of the solutions a word which was to be used in the experiment, the context was accepted as strongly-constraining. If none of the judges suggested the appropriate word, the context was considered weakly-constraining. The sentence "The university choir performed a song at the inauguration of the year" may serve as an example of a sentence in which the intended word "song" is used in a strongly-constraining context. The very same word in a weakly-constraining context is as follows: "From behind the door a song performed by the university choir was heard" (all used words and contexts have been presented in the Appendix 1).

Target words were never presented in the sentence-ending position to avoid interference with the semantic and syntactic integration processes in a sentence.

All words and sentences used in the experiment are included in Appendix 1.

Subjects

32 persons participated in the experiment – 16 women and 16 men, aged between 19 and 30. All the subjects were students in the Faculty of Psychology of the University of Warsaw.

Experimental procedure

The subject's task was to read sentences presented on the computer screen. The words of each sentence appeared consecutively, with previous words disappearing once a new one appeared. The spatial position of words in a written sentence was preserved. The "self-paced reading" procedure was applied in the experiment: the subject caused a next word to appear by pressing a spacebar on the keyboard. If the spacebar was not pressed for 4000 ms, the next word was automatically shown. In each sentence presented to a subject there was one word with scrambled letters.

The subjects were asked to perform the task as quickly as possible, but to advance to the next word only when they were sure that they had understood the presented word (the instruction announced to the subjects is included in Appendix 2.). In order to assure that subjects complied with the instructions, and did not press the spacebar automatically, they were told that they would be asked to paraphrase some of the sentences. The experimenter asked the subject to paraphrase approximately 25% of the sentences, selected at random.

Experimental design

The subjects were divided into two groups. Twenty sentences were presented to each of them. Each group received 10 sentences with frequent target words and 10 with infrequent target words. In order to avoid showing the same sentence or the same scrambled word to a subject twice, five frequent words appeared in a strongly-constraining context and five in a weakly-constraining context. The same rule applied to infrequent words.

Three practice sentences preceded 20 test sentences.

The reading time of words with scrambled letters was measured. It was understood as the interval between the word's appearance and the moment when the subject pressed the spacebar advancing to a next word.

The experiment was programmed using E-Prime software.

Statistical analysis

Four results of 4000 ms were removed from the analysis (0.6% of all the responses). These results occurred because the subject had waited until the program projected the next word without giving any response. The removed results were replaced with

Table 1. Averages (standard deviations were given in brackets)

CONTEXT	FREQUENCY		
	RARE WORDS	FREQUENT WORDS	
WEAKLY-CONSTRAINING	743.509 (290.930)	647.556 (225.115)	**695.533** (279.073)
STRONGLY-CONSTRAINING	655.932 (203.430)	619.293 (178.505)	**637.613** (190.745)
	699.721 (252.902)	**633.425** (202.035)	

an average, calculated from the mean result of a given subject and a given word.

As described above, five sentences from each experimental condition were shown to each subject. They consisted of 5 sentences including frequent words in strongly-constraining contexts, 5 sentences with frequent words in weakly-constraining contexts, 5 infrequent words in strongly-constraining contexts, and 5 in weakly-constraining ones. All words belonging to a given category were treated as equivalent. Thus, before the actual analysis of the results was commenced, four averages for each subject had been calculated; each average concerned the five words belonging to a given category. In this way, for every subject four results were calculated.

The data were subject to the ANOVA for repeated measures. The intra-object factor was the frequency of words and the context of their occurrence. As inter-object factor the group to which the subject belonged was added.
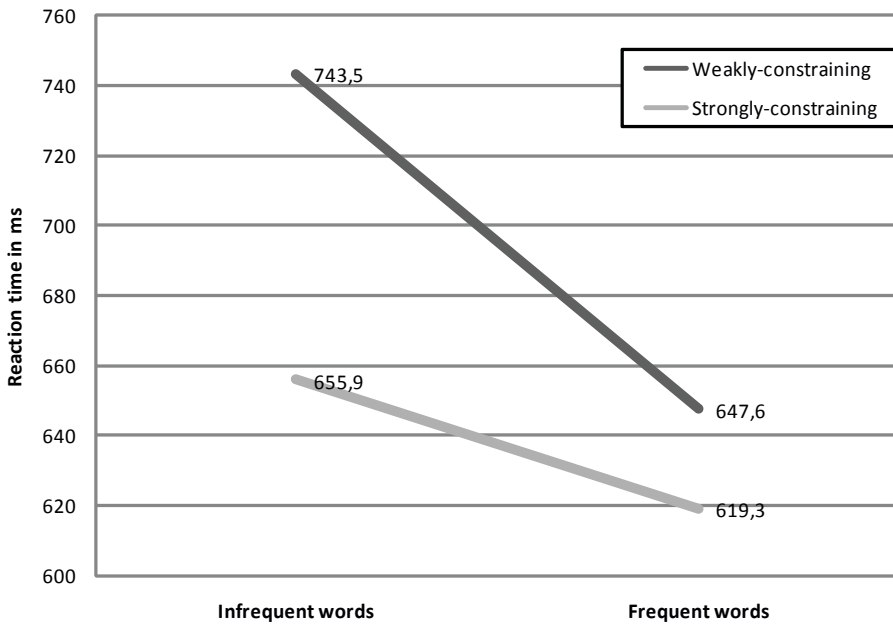
Results

Means and standard deviations obtained are presented in Table 1 and in Figure 1.

The frequency*context interaction effect was not significant (1, 30) = 2.628, p = 0.115. The significant main effect of frequency F(1, 30) = 9.834, p = 0.004 has been found, as well as the significant effect of context F(1, 30) = 8.819, p = 0.006. The infrequent words were thus recognized more slowly than the frequent words and the words in a strongly constraining context were recognized more easily than the words occurring in a weakly constraining context.

Discussion

The main goal of this study was to demonstrate that attempts to explain the "jumbled words" effect cannot be undertaken without turning to knowledge con-

Figure 1. The influence of context and frequency of words on the time of word recognition

cerning the "top-down" processing. The results confirm also the importance of the "top-down" processing in word recognition in general.

In this study two operational hypotheses were formulated, concerning the influence of context in which the words occur and words' frequency in the language on the recognition of words with scrambled letters. Both hypotheses have been confirmed. The experiment demonstrated that the "jumbled words" created from words occurring infrequently in the language are recognized more slowly than "jumbled words" created from frequent ones and that the "jumbled words" presented in a strongly-constraining context are recognized more easily than "jumbled words" presented in a weakly-constraining context.

It must also be noted that, even though the interaction effect of context and frequency factors was not statistically significant, Graph 1 may suggest that the thesis of McClelland and Rumelhart (1981) was not unjustified. Their view assumed that if the context plays an important role (that is, when it is strongly-constraining), the role of frequency of the word recognition should be smaller.

In light of the above results, it seems reasonable to conclude that models aspiring to the explanation of "jumbled words" recognition should take both the "bottom-up" and the "top-down" factors into account. That includes not only the influence of the word frequency, but also the semantic context in which the words appear.

It remains to be determined in further research if the syntactic and situational contexts also play a role, but, on the basis of this study one could expect that the answer would be positive.

Bibliography

Coltheart, M. Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.

Fodor, J. A. (1983). *Modularity of mind: an essay on faculty psychology.* Cambridge, Mass.: MIT Press.

Gleason, J. B. & Ratner, N. B. (Ed.) (1998; Polish edition: 2005). *Psycholingwistyka.* Gdańsk: GWP.

Gough, P. B., Alford, J. A., & Holley-Wilcox, P. (1981). Words and contexts. In O. J. L. Tzeng & H. Singer (Eds.), *Perception of print. Reading research in experimental psychology* (pp. 85-102). Hillsdale, NJ: Erlbaum.

Grainger, J. (1990). Word Frequency and Neighborhood Frequency Effects in Lexical Decision and Naming. *Journal of Memory and Language*, 29, 228-244.

Grainger, J. & Whitney, C. (2004) Does the huamn mnid raed wrods as a wlohe?, *Trends in Cognitive Science,* 8, 58-59.

Howes, D. H. & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. Journal of Experimental Psychology, 41, 401-410.

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, 22, 517-560.

Korpus języka polskiego PWN.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-405.

Miller, G. A. (1951*) Language and communication.* New York: McGraw-Hill.

Simpson, G. B., Peterson, R. R., Casteel, M. A., & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 88-97.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 20*, 120-136.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27,* 324-340.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143-154.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. *Psychonomic Buletin and Review, 8,* 221-243.

Velan, H., & Frost, R (2007). Cambridge University versus Hebrew University: The impact of letter transposition on Reading English and Hebrew. *Psychonomic Bulletin and Review, 14* (5), 913-918.

Appendixes

Appendix 1

All words and contexts matched to them in the experiment (below they are given both in Polish and in English language version). The frequency of word occurrence has been indicated in brackets – in accordance to *Korpus języka polskiego PWN*. The word with scrambled letters has also been given.

RARE WORDS – STRONGLY-CONSTRAINING CONTEXT

hobby *hobby*	hboby	(6)	Modelarstwo było jego hobby i pasją od dzieciństwa. *Modeling has been his hobby since childhood.*[1]
poród *delivery*	poórd	(10)	Położna odebrała poród w zastępstwie lekarza. *The midwife received the delivery instead of the physician*
klapa *lid*	kplaa	(11)	Opadająca ciężko fortepianowa klapa przytrzasnęła Jankowi palce. *Heavily falling piano lid crushed John's fingers.*
bukiet 11 *bunch*	bkuiet	(11)	Kamil kupił w kwiaciarni duży bukiet czerwonych róż. *In the flower shop, Kamil bought a large bunch of red roses.*
bufet *canteen*	bfuet	(16)	Pierogi serwowane przez wydziałowy bufet były dla Marcina za drogie. *Ravioli served by the university canteen were too expensive for Marcin.*
sonda *probe*	sndoa	(8)	Na Marsa wysłana została sonda bezzałogowa. *An unmanned probe has been sent to Mars.*
deser *dessert*	dseer	(10)	Po obiedzie Janek miał jeszcze ochotę na deser lodowy. *After dinner Johnny still felt like eating an ice-cream dessert.*
stróż *wachman*	sótrż	(9)	Bloku Pawła pilnował siedzący w budce stróż wynajęty przez mieszkańców. *Paul's apartment building was guarded by a watchman hired by the residents.*
wanna *bathtub*	wnana	(17)	W łazience stała biała wanna i pralka. *In the bathroom there was a white bathtub and a washing-machine.*
kubek *mug*	kbuek	(19)	Spragniony Andrzej sięgnął do zmywarki po kubek i nalał do niego wody. *Andrew was thirsty, so he reached into the dishwasher for a mug and poured water into it.*

---

[1] The English translations of the sentences do not always convey the manner in which the context narrowed down the possibilities. This is due to the fact that Polish has a much more unconstrained word order than English.

RARE WORDS – WEAKLY-CONSTRAINING CONTEXT

hobby	hboby	Dla Andrzeja najważniejsze było jego hobby i poświęcał mu wiele czasu.
		For Andrew, his hobby was most important and he devoted a lot of time to it.
poród	poórd	Jurka stresował zbliżający się poród żony.
		George was under stress because of his wife's approaching delivery.
klapa	kplaa	Kamilowi przycięła palce klapa od fortepianu.
		Kamil's fingers were crushed by the lid of the piano.
bukiet	bkuiet	Janek kupił swojej narzeczonej bukiet czerwonych róż.
		Johnny bought his fiancée a bunch of red roses.
bufet	bfuet	W firmie Marcina był bufet i sklepik.
		There was a canteen and a shop in Marcin's firm.
sonda	sndoa	Wysłana przez rządy obu krajów sonda dotarła już na Marsa.
		A probe sent by the governments of the two countries has already reached Mars.
deser	dseer	Maciek zjadł na deser rurkę z kremem i pączka.
		For desert, Maciek ate a cream-wafer and a doughnut.
stróż	sótrż	Wychodząc rano Michał zauważył, że stróż pilnujący domu zasnął.
		On leaving in the morning, Michael noticed that the watchman guarding the block had fallen asleep.
wanna	wnana	Na wystawie stała biała wanna i komplet mebli łazienkowych.
		There was a white bathtub and a set of bathroom furniture on display.
kubek	kbuek	Jurek wyjął z szafki kubek i zrobił sobie herbatę.
		George took the mug out of the cupboard and made himself some tea.

FREQUENT WORDS - STRONGLY-CONSTRAINING CONTEXT

hasło *word in* *cross-word* *puzzle*	hłsao	(91)	Rozwiązaniem krzyżówki było hasło pięcioliterowe. *The solution in a crossword puzzle was a five letter word.*
pieśń *song*	pśień	(93)	Chór uniwersytecki wykonał pieśń na inauguracji roku. *The university choir performed a song at the inauguration* *of the year.*
obiad *dinner*	obaid	(84)	Po szkole Dawid przygotował obiad dla siebie i siostry. *After school David prepared dinner for himself and his* *sister.*
mistrz *Champion*	mtsirz	(90)	Brązowy medal otrzymał ubiegłoroczny mistrz w tej dys- cyplinie sportu. *The bronze medal went to last year's champion in this* *discipline.*
termin *term*	terimn	(139)	Wyznaczono październik jako termin przyspieszonych wyborów. *October was chosen as the term for accelerated elections.*
uczeń *Pupil*	ucezń	(113)	Najlepszy w klasie uczeń otrzymał ocenę celującą. *The best pupil in class received the full mark.*
wiara *faith*	waira	(160)	Katolików charakteryzuje niezachwiana wiara w nieśmier- telność duszy. *Catholics are characterized by an unshaken faith in the* *immortality of the soul.*
model *model*	moedl	(160)	W salonie stał najnowszy model samochodu. *In the showroom there was the latest model of a car.*
sklep *shop*	seklp	(116)	Po drodze Arek odwiedził sklep sportowy. *On his way, Arek visited a sport shop.*
obszar *terrain*	osbzar	(173)	Lasy pokrywają cały obszar kraju. *Woods cover the whole area of the country.*

FREQUENT WORDS – WEAKLY-CONSTRAINING CONTEXT		
hasło	hłsao	Andrzej po zastanowieniu wpisał hasło do krzyżówki.
		After consideration, Andrew entered the word into the crossword puzzle.
pieśń	pśień	Zza drzwi słychać było pieśń wykonywaną przez chór uniwersytecki.
		From behind the door a song performed by the university choir was heard.
obiad	obaid	Marek umówił się z Kasią na obiad w restauracji.
		Mark made an appointment with Kate for dinner in a restaurant.
mistrz	mtsirz	Staś był znany jako mistrz w swoim fachu.
		Staś was known as a champion in his trade.
termin	terimn	Grześka przestraszył zbliżający się termin oddania projektu.
		Greg was scared by the approaching dead-line for delivering the project (dead-line is another equivalent for termin *in Polish).*
uczeń	uczń	Annę odwiedził jej dawny uczeń z klasy humanistycznej.
		Ann was visited by her old pupil from the humanities class.
wiara	waira	Dla Michała najważniejsza była wiara i ojczyzna.
		For Michael faith and homeland was of greatest importance.
model	moedl	Antkowi podobał się ten model samochodu.
		Antek liked this model of a car.
sklep	seklp	Po namyśle Emilia postanowiła otworzyć sklep obuwniczy.
		After consideration Ann decided to open a shoe-store.
obszar	osbzar	Muzyka ludowa to obszar zainteresowań Magdy.
		Folk music is Magda's field of interest.

Appendix 2

Instruction used in the experiment

Words constituting sentences will appear on the screen. Your task is to read those words and to press the "space" key after reading them. **In some words certain errors (misprints) may occur. My goal is to check if you are able to read them nonetheless**.

Pressing the "space" key will move you to the next word. When the next word appears the previous one will disappear – you should thus try to read them carefully because you will not be able to go back to the previous one. **It is important that you should press the "space" key only if you are sure that you have understood the given word**. From time to time I shall ask you to repeat the sentence you have just read.

If it happens that you are not able to read a word, wait until the computer passes to the next one.

At the beginning you will receive three trial-sentences so that you can understand in practice what this task is about. Only later we will pass to the test itself.

If you have any doubts please tell about them now.

Remember! Understanding words is important. Press the "space" key as quickly as possible but only after you have read and understood each word.